DOCUMENT RESUME

ED 359 027 SE 053 152

AUTHOR Kjoernsli, Marit; Jorde, Doris
TITLE Evaluation in Science: Content or Process?
PUB DATE Apr 92
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April, 1992).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Evaluation Methods; Foreign Countries; *International Studies; Knowledge Level; *Multiple Choice Tests; Pilot Projects; Science Education; Science Tests; Student Evaluation; *Test Construction; Test Validity
IDENTIFIERS Assessment of Performance Unit (United Kingdom); *International Assn Evaluation Educ Achievement; Open Ended Questions; Performance Based Evaluation; Science Achievement; Science Process Skills; *Third International Math and Science Study

ABSTRACT

Science assessment has been included with mathematics and language assessment on the international level since the 1970s. This paper discusses techniques of assessment that have been utilized to measure science process skills. The first section discusses the Assessment Performance Unit, a British project with the aim of developing innovative methods in assessing science achievement. Six categories of science activities for assessment purposes were identified by the project: (1) use of graphical and symbolic representation; (2) use of apparatus and measuring instruments; (3) observation; (4) interpretation and application; (5) planning of investigations; and (6) performance of investigations. The second section discusses the International Science Studies conducted by the International Association for Evaluation of Educational Achievement (IEA). Assessment items used for the first and second studies are described. The remainder of the paper presents the assessment techniques that were piloted for use in the third study. The following categories of questions were used in the Third International Math and Science Study: (1) multiple choice items; (2) open-ended written items; (3) performance tasks, which produce a physical product beyond writing; and (4) performance tasks where the process of actually doing the task is documented and examined. Sample items for each of these categories are presented and discussed. (Contains 16 references.) (MDH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED359027

A paper for the American Educational Research Association 1992 Annual Meeting

Evaluation in Science:

Content or Process?

Marit Kjærnsli and Doris Jorde
Centre for Science Education
University of Oslo
Norway

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Marit Kjaernsli

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Introduction

The purposes of educational evaluation/assessment generally have been manifold: Diagnosing, counselling, selecting, certifying, and evaluation of curricula, teaching, or educational systems (Johnson, 1987). Compared to such subjects as language and mathematics, science has a shorter history of widespread activity in the field of assessment. It is only since the 1970's that science has been included in national and international assessment programs.

Early assessment in science concentrated on testing factual knowledge in the sciences. Questions were asked which typically had a "correct" answer. The growing knowledge of how children learn science led to an increasing change in the goals and objectives of teaching science. Science teaching has progressed towards an activity based, process orientated curriculum - a change that can not always be reflected in quantitative assessment instruments. The result is that innovation in science assessment has often fallen behind innovation in science teaching.

It is imperative that good instruments be developed to assess students' understanding in science. The instruments will need to assess not only factual information but also the manner in which we go about doing and learning science. Science assessment is essential for providing information on misunderstandings and alternative conceptions, as well as the possible reasons for why such obstacles occur.

Science assessment includes large national and international tests which are used for international comparisons. Science assessment also includes that which teachers do in their classrooms on a regular basis. Though inter-related in many ways, the main focus of this paper will be on the large international science assessment projects conducted by The International Association for Evaluation of Educational Achievement (IEA).

SE053152

The processes of Science

In the beginning of the 1970's there was a growing interest in the processes of science in science teaching. Not only was it important to learn factual information in science, but equally important was the way one went about learning science. The processes we refer to here include: observation, hypothesis testing, experimentation, classification and communication. Science curriculum materials were developed with an emphasis on processes including: Science A Process Approach (SAPA, 1967); Science Curriculum Improvement Study (SCIS, 1974); The Nuffield Project and Science 5/3 (1972).

Science process skills generally divide between those that are cognitive in nature and those that relate to practical activities. Manipulative and observational skills, for example, belong to the latter category, whilst recall and application of knowledge, the interpretation of information and problem-solving are examples of cognitive skills. It should be noted that the distinction between cognitive and practical skills is frequently only a matter of convince, for in many actual situations encountered in science education they come together. For example, being able to follow instructions accurately for conducting experiments, may be a skill that relates primarily to the execution of a practical task, but also invokes a significant cognitive element (Kempa,1986).

It seems natural that if process oriented objectives and activities are emphasized in curricula, they should also be focused on in the assessment methods. However, science assessment tends to lag behind science teaching objectives. Testing and assessment methods are, in fact, often in conflict with the objectives expressed in the curriculum (Angell and Lie, 1990; Horsfjord and Dalin, 1988; Johnson, 1987; Raaen, 1990; Swan, 1991). Until the end of the seventies, the main part of all tests in science asked for a mere reproduction of factual information, even though Bloom's taxonomy of cognitive objectives was often used as the basis for the assessment and test specification (Bloom, 1956).

The Assessment Performance Unit (APU)

One of the first assessment projects that concentrated on the processes of science took place in England beginning in 1975. The Assessment Performance Unit (APU) project had the aim of developing innovative methods in assessing science achievement in both processes and content for pupils 11, 13 and 15 years old.

The following APU framework for assessment reflects the underlying view of science adopted. Six categories of science activities were identified for assessment purposes. The framework is common to all three age groups.

1 *Use of graphical and symbolic representation*
- *reading information from graphs, tables and charts*
- *representing information as graphs, tables and charts*

2 *Use of apparatus and measuring instruments*
- *using measuring instruments*

- *estimating physical quantities*
- *following instructions for practical work*

3 *Observation*
- *making and interpreting observations*

4 *Interpretation and application*
- *I interpreting of presented information*
- *II applying: Biology concepts, Physics concepts, Chemistry concepts*

5 *Planning of investigations*
- *planning parts of investigations*
- *planning entire investigations*

6 *Performance of investigations*
- *performing entire investigations*

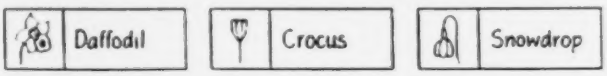
The investigation surveys consist of written, individual practical and group practical tests, all designed to assess how children "do" science by using the processes of science.

The APU researchers point out that there are skills and processes in science that only can be assessed by practical tests. In APU it was shown that valid and reliable assessment can be made of the complex activities that practical tests imply. The drawback of such testing is that it is resource demanding, both with time and equipment. However, the developing of methods for assessing and categorizing science process and content skills may form a valuable basis for diagnostic questions and tests that teachers can use in their own pupil assessment practices. The assessment of processes together with content in science allows teachers to gain incites into pupils thinking and reasoning.

The following (item 1) is an example of a paper and pencil test from category 4; Interpretation and application. It is also a typical example of how APU went about developing a series of ingenious questions and test items to assess processes. Their particular feature is that many of these explore "everyday" situations and do not therefore require the pupil to possess specialized scientific knowledge.

Item 1

Mr. Brown had a garden full of daffodils, crocuses and snowdrops, which came up year after year.



For three years Mr. Brown kept a record of when the plants where in flower. This is what they looked like.

	EARLY JAN	LATE JAN	EARLY FEB	LATE FEB	EARLY MARCH	LATE MARCH	EARLY APRIL	LATE APRIL	EARLY MAY
YEAR 1									
YEAR 2									
YEAR 3									

(Mr. Brown forgot to put snowdrops on the record in year 3!)

(a) What pattern do you notice in the chart about the times at which crocuses and daffodils flowered?

.....

.....

(b) When do you think the snowdrops were in flower during year 3?

.....

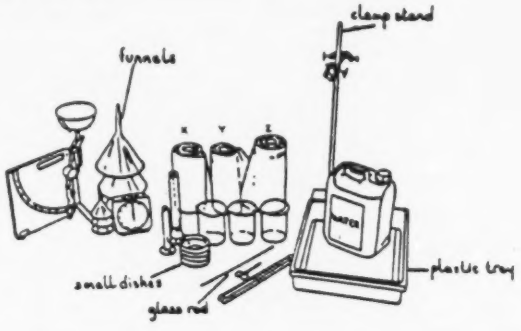
Item 2, the "Paper towel test" is an example of category 6; Performing of investigations. Questions in this category consist of relatively open-ended practical tasks in which pupils are given 30 minutes to carry out experiments, using some or all of a number of items or apparatus, to solve a particular problem. The task is presented in a standardized way by a trained administrator who then records, on a checklist, details of pupils' experimental technique, results and conclusions.

Item 2

You have in front of you three kinds of paper towel X, Y and Z. This is what you find out:

Which kind of paper will hold most water?

You can use any of the things in front of you. Choose whatever you need to answer the question.
Make a clear record of your results, so that I can understand what you have found out.



One quickly notices that the examples from APU tasks are very different from typical multiple-choice questions or questions that assume a correct answer. Children are asked to use the processes of science in their solutions and there may be multiple solutions to a problem.

The APU has influenced science assessment at every level. At the classroom level it has provided tools for activity based assessment of content and processes. At the level of national and international science assessment, APU has guided the way for innovative ideas in science process assessment using quantitative instruments.

The International Association for Evaluation of Educational Achievement (IEA)

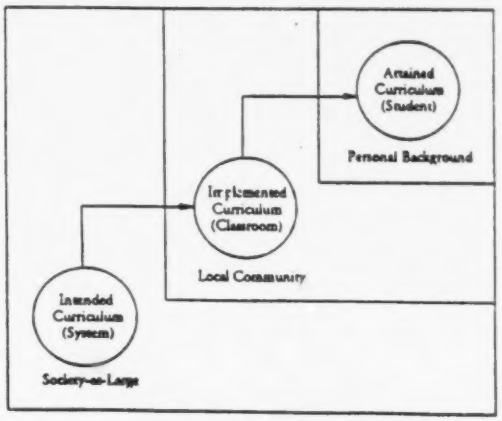
The IEA was created in the beginning of 1960 as an international association of research centers. Countries decide themselves whether or not they wish to participate in the association, and today there are over 50 countries in the membership. IEA conducts international studies in different subject areas with the following aims:

1. *Assess the potential impact that alternative curricular, teaching, and administrative strategies have on student achievement within countries.*
2. *Provide current international information which countries can use to compare and contrast their curricula, teaching practices, and student outcomes with those from other countries of interest.*

IEA has evaluated science achievement internationally since 1970, beginning with the First International Science Study (FISS) project. The Second International Science Study (SISS) test was undertaken in 1984, and The Third International Mathematics and Science Study (TIMSS) will be administered in 1994.

The aim of the science achievement tests is to compare the intended and implemented curriculum to the attained (or what children actually learn). The figure below illustrates the complexity of the educational environment.

Figure 1. Conceptual Framework for TIMSS



IEA science tests are large scale psychometric tests which are dominated by multiple-choice format questions. This format greatly constrains the type of question one may ask as compared to the APU open question format. There have been attempts, however, in the development of these tests to experiment with alternative forms of assessment so that the test becomes a better picture of how science is actually taught in schools.

The First International Science Study (FISS) was conducted in 1970 with 17 participating countries. In addition to the main multiple-choice format test which was completed by all countries, England and Japan administered a practical test at the ninth grade level. The test consisted of 5 separate tasks. The results of the practical test were compared to the results of a "paper-and-pencil practical test" and a "paper-and-pencil-achievement test". Comber and Keeves (1973) concluded that the practical test could measure skills that were different from what the usual written test was suited for. Also, these skills could only to a certain degree be tested through the written test where practical problems were presented ("paper-and-pencil practical tests").

The Second International Science Study (SISS), conducted in 1983 included a study of science achievement in 24 countries at three levels in each school system: age 10, age 14 and the final year of secondary school. This study was a follow up to the FISS study and built on results from FISS. The curriculum grid developed for the first study drew upon, to a large extent, the significant work on evaluation in the field of science prepared by Bloom and his colleagues and was further refined for employment in this study.

There had been two strong criticisms of the FISS test. First, some items were too "wordy", so that it was not clear whether reading skill was being measured or whether competence in science was being assessed. Secondly, the items that had been employed as pencil and paper tasks to assess achievement with respect to practical work in science, did not measure in a valid way process skills in science, nor did they measure science competence as effectively as did the other content based items. These two shortcomings suggested that greater use should be made of diagrammatic material in the framing of questions, and that all questions should be more closely related to the body of scientific content taught in schools, and that pencil and paper practical items should not be incorporated as a specific subset of the test items (Rosier and Keeves, 1992).

In the SISS investigation, six countries (Israel, Japan, Korea, Singapore, Hungary and USA) included a practical test for ages 10 and 14. The skills were categorized as follows: performing, investigation, and reasoning. Preliminary results indicate little correlation between the actual SISS science test and the experimental practical science test. Tamir (1987) stresses the importance of international comparisons of this type which help to illuminiate differences between the intended and implemented curriculum between countries.

The Third International Math and Science Study (TIMSS)

For the first time in IEA traditions, Mathematics and Science will be combined into one international project. Currently 70 countries have expressed an interest in participating in this project which will be administered during the 1993-94 school year. As compared with FISS and SISS, the test populations have been changed to include ages 9, 13 and the last year of secondary school.

TIMSS builds on the results of previous IEA science and math studies. IEA studies are designed to address a broad spectrum of questions and issues of interest in particular field of discipline. Thus, SIMS and SISS addressed many issues which were of paramount importance to mathematics and science educators in the 1980's. TIMSS will continue this pattern addressing issues such as:

* *international variations in the mathematics and science curricula;*
* *opportunity to learn;*
* *attitudes and opinions of students teachers;*
* *students' achievement, with particular emphasis on capability of students to apply their knowledge and skill in non-routine applications;*
* *the role of technology in teaching and learning of mathematics and science;*
* *participation rates in college preparatory courses in mathematics and science, with particular regard to gender-based differences in rates of participation;*
* *practices employed by schools and school systems to direct students' coarse selection, including tracking and streaming;*
* *the nature, role, and influence of officially prescribed textbooks on the teaching of science and mathematics;*
* *the comparative efficacy if different approaches to the teaching of mathematics and science on student outcomes.*

The measurement techniques to be used in the TIMSS project must be reflective of the

current educational goals within science and mathematics education, including students' reasoning, problem-solving and communicating techniques. Traditional multiple-choice format will thus be supplemented by alternative and innovative assessment techniques.

The current framework for assessment in the TIMSS project is as follows:

1 *Traditional multiple choice items*

2 *Open-ended written items which require both short answers and longer, essay type responses*

3 *Performance tasks which produce a physical product beyond writing*

4 *Performance tasks where the process of actually doing the task is documented and examined*

These alternative assessment techniques will be useful in testing parts of the implemented curriculum that previously were not possible to assess due to testing constraints. Items will allow students to supply answers rather than just selecting answers to questions. In addition, an emphasis will be placed on alternative answers ("incorrect") to questions which may help to illuminate student alternative conceptions.

A pre-pilot test was administered in most countries participating in the TIMSS project in the fall of 1991. The goals of the Pre-pilot Testing were to introduce the National Project Coordinators to the problems associated with translation, relations with schools, data preparation, communication of data to the test center, and other associated problems.

The following discussion is based on the pre-pilot test items which demonstrate the categories of question types used in the TIMSS framework for assessment. Item examples are taken from the test given to 13 year olds. Comments on the results obtained from the individual items are taken from a joint TIMSS Report (Brekke, Kjærnsli, Lie et.al, 1992) based on the pre-pilot experiences of Norway, Sweden and Denmark.

Multiple - choice items

As discussed previously, the most common type of question found in IEA testing is the multiple choice format. This type of question has most often been used to get at factual information where students select the correct answer and the alternative answers are incorrect. Item 3 is an example of such a question taken from the TIMSS Pre-pilot test.

Item 3

When sand is thrown onto a fire, it puts the fire out by cutting off the supply of

A. oxygen
B. nitrogen
C. helium
D. carbon dioxide

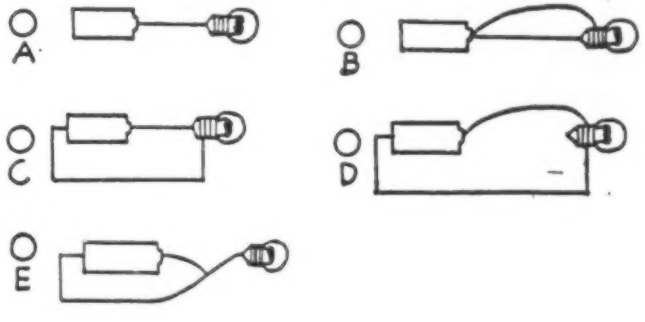
The over-all aim of the multiple-choice item is seemingly to distinguish between "right" and "wrong" answers. However, we strongly feel that the diagnostic perspective should also be taken into account. By this we mean that TIMSS offers a unique possibility to explore pupils' thinking within different subject topics. And in this regard wrong answers give more interesting information than the correct ones. The choice of distractors is therefore very important.

There has been a substantial emphasis in the last few years in the field of student learning, especially as related to studnet "Alternative Conceptions" or "Misconceptions" in science and mathematics. We therefore have a well documented source of references from which we can construct good "distractors" in order to get important information about student thinking.

Item 4 demonstrates a question which has carefully chosen alternative answers taken from the alternative conceptions commonly found amoung students in the area of electricity (EKNA, 1979-1989).

Item 4

The bulb in the figure below is connected to a battery with a wire. In what figure will the bulb light up?
Put a tick in the correct answer.



Research on pupils' alternative conceptions in electricity shows that a lot of students have "one-pole" understanding (misconception) of the bulb and "one-pole" understanding (misconception) of the battery. In Item 4 the distractors are made according to this research. Distractor a: "one-pole" battery and "one-pole" bulb; distractor b: "one-pole" battery and "two-pole" bulb; distractor d: "two-pole" battery and "one-pole" bulb; distractor e: "two-pole" battery and "one-pole" bulb.

This type of question clearly gives diagnostic information on how children are thinking about electricity.

Multiple choice items have also been used to test the processes of science. These items tend to put students into situations where they must think through a problem and find the best possible answer. Because the process information is so demanding, the content information in these questions if often very elementary as demonstrated by Item 5.

Item 5

A student wanted to learn which of three types of soil (clay, sand, or loam) would be best for growing beans. Three flower pots were filled, each with a different type of soil. The same number of beans were then planted in each. The pots were placed side by side on a window sill and each pot given the same amount of water.

The drawing shows the pots and the results after a few days.

Why was the experiment NOT a good one for the purpose?

A. The size of the pots was not the same
B. One pot should have been placed in the dark.
C. The plants would get too hot on the window sill.
D. Different amounts of water should have been used.

It is, in fact, not necessary to know any content matter in order to answer this process question. The question tests "control of variable". Multiple choice questions test either content or process, but almost never both at the same time. When content and process are combined in the same question, students answering correctly are assumed to have managed both content and process elements in the question. However, it is unclear if the process or the content element has been misunderstood by those students who choose the incorrect answer.

Open-ended written items

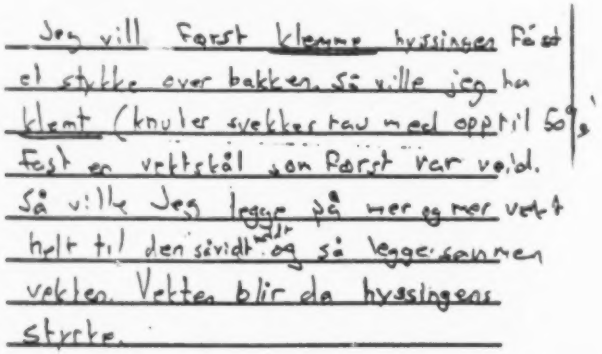
Open-ended items differ from locked answer test items in that there is no one correct answer to the problem given. Students are given the opportunity to provide a variety of arguments which in turn may lead to variation in the solution to the problem. As mentioned earlier in the paper, APU has played a substantial role in the introduction of open-ended format questions in science; a subject that is typically assessed by questions assuming a "correct" answer.

Item 6 is an example of an open-ended question taken from the TIMSS pre-pilot test.

> Item 6
>
> You have a piece of string and you want to know how strong it is. Write down what you think might be the best way to test the strength of your piece of string.
>
> _____
>
> _____

The two following examples demonstrate typical student answers to Item 6; answer 1 coming from the category "complete answer", answer 2 coming from the category "average answer". It is easy to see how this type of item allows us to understand how students are thinking when they are solving problems. In this way both content and process information may be assessed in the same question. Problems with open-ended questions are often related to administration and interpretation in that they are time consuming and difficult to objectively code.

Answer 1:


Jeg vill først klemme hyssingen på
et stykke over bakken. Så ville jeg
klemt (knuter svekker tau med opptil 50%)
fast en vektskål som først var veid.
Så ville Jeg legge på mer og mer vekt
helt til den sprakk. og så legge sammen
vekten. Vekten blir da hyssingens
styrke.

I would first fasten the string a little bit above the ground. Then I would fasten (the knot weakens the string up to 50%) a scale plate that first was weighed. Then I would add more and more wight until it just held and then I would add up the wight. The weight is then the string's strength.

BEST COPY AVAILABLE

Answer 2:

Man kan binde bysnóren fast i vekter og se når den tyker.
Dermed ser man på antall vekter den tålte hvor sterk bysnóren er.

One can tie the string to a weight and see when it breaks. Then you see how much weight it held (and then you see) how strong the string is.

In large scale IEA testing, open-ended testing has been tried on an experimental basis. The TIMSS test will try to incorperate this type of item, though not without difficulty. The pre-pilot test information has clearly demonstrated that if such questions are to be included, extensive testing of the items must be done beforehand such that detailed information may be provided for how each item is to be coded. Included in this information should be examples of typical student answers from different categories of the coding.

Performance tasks

Performance tasks assume that students are given a practical problem to solve. They are often characterized by the introduction of equipment as a part of the problem solving activity. A written account of the process of solving the problem is most often required at the completion of the task. Many of the APU test items fall into this category as represented by Item 2; The Paper Towel Test.

Item 7 is an example of a performance task used on the TIMSS pre-pilot test for all three populations, where students were asked to work with a partner.

Before solving the performance item, the pupils were given the following directions:

This item is actually a problem-solving activity involving both mathematics and science. As in any scientific investigation, you may have to make a plan, execute it, and then record your plan, actions, and results.
You will be working with a partner for this activity. you may quietly discuss your thoughts and plans with your partner and may work together towards the solution of the problem. However, each of you must write your own answers in your own test booklet. You have 15 minutes to complete this activity.

BEST COPY AVAILABLE

Item 7

You are presently working at a desk, table or some other piece of furniture. What size is it?

The following examples are considered "average" student responses.

Answer 3.

Vi måtte sidene på pulten med arkene til prøven vi fant ut at lengden var 70 cm og bredden var 55 cm. Arket er 30 cm langt i lengden
O: 70+70+55+55= 250 cm
A: 70·55= 3850 cm²

We measured the sides of the desk with the test paper. We found out that the length was 70 cm and the width was 55 cm. The paper is 30 cm long in length.
Perimeter: 70+70+55+55=250cm
Area: 70 55=3850cm

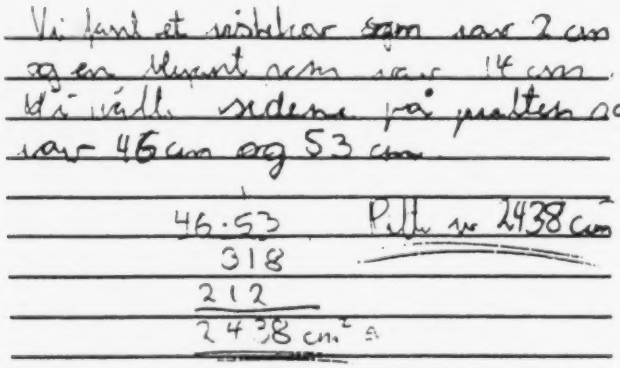
Answer 4.

Jeg prøver å tenke meg hvor mye 1cm er så setter jeg opp streker på pulten og teller dem oppetter på i lengden er ca 75cm og Bredden er ca 56 cm. 75cm·56cm = 4200cm²
Omkretsen er 75cm+75cm+56cm+56cm = 262cm

I try to think to myself about how much 1cm is and then I set up marks on the desk and add them up afterwards. The length is about 75cm and width is about 56cm. 75cm 56cm=4200cm. The perimeter is 75cm+75cm+56cm+56cm=262cm

BEST COPY AVAILABLE

Answer 5.


Vi fant et viskelær som var 2 cm
og en blyant som var 14 cm.
Vi målte sidene på pulten som
var 46 cm og 53 cm.
46 · 53
318
212
2438 cm²
Pulten var 2438 cm²

We got an eraser that was 2cm and a pencil that was 14cm. We measured the side of the desk as 46cm and 53cm

It is easy to see from these examples that there is no one "correct" answer to the problem. Students develop a plan for solving the problem and then proceed to follow through on the plan. After they have obtained results they are asked to write down the proceedure they have completed including the data.

When practical items are done correctly, assuming enough time and information, students are able to work actively with the processes of science while at the same time solving a problem. This type of item is the best for representing the overall goals of the science lesson. However, in large scale testing, this type of item is not without complications.

The problems associated with coding for performance task items are the same as those mentioned for open-ended items. Items must be pre-tested and categories established for coding before the actual test is given.

Performance items are not familiar ways of assessing students in science and mathematics, therefore some time is needed to introduce the proceedure. In addition, when many small groups of students preform the task at the same time, they often look around to see what others are doing and then "steal" ideas from each other. This problem may be alleviated if simple equipment is given to each student rather than asking students to work in groups.

BEST COPY AVAILABLE

distractors which provide diagnostic evaluation. We have also seen the emergence of alternative assessment methods in science which include open-ended questions and practical tasks. These newer methods have many strengths in the information provided, however they take time to administer and code, making them difficult to justify in large projects.

The following chart is a short summary of the assessment methods discussed in this paper, their strengths and their weaknesses.

Figure 2. Science and Mathematics Assessment Methods in TIMSS

Assessment methods	Content/process Positive:	Negative:	Objectivity/administration etc Strengths:	Weaknesses:
Multiple-choice	* large content coverage * different cognitive levels may be tested * diagnostic evaluation possible	* either content or process tested, rarely both * reflective, thinking process absent	* objective, high reliability * easy coding and administration * easy to assess large populations	* difficult to construct good distractors
Open-ended	* complex answers possible: both process and content * relective, thinking process possible to show * depth in content coverage	* limited content coverage	* high validity possible * items easy to construct	* time consuming to code/mark * objectivity difficult to achieve in large populations * reliability difficult to achieve in large populations * difficult to assess large populations
Performance task	* process skills easily tested * group assessment possible * complex answers possible: both process and content * reflective, thinking process possible to show * depth in content coverage	* limited content coverage	* items easy to construct * high validity possible	* requires equipment * time consuming to administer * difficult to assess large populations * objectivity difficult to achieve in large populations * reliability difficult to achieve in large populations

Multiple-choice questions will continue to dominate large international science and mathematics assessment projects. In addition to their usefulness in testing for factual knowledge, better construction is now making it possible to use multiple-choice questions for diagnostic evaluation of student understanding.

Open ended questions, including performance tasks, provide testing options which more closely relate to the science teaching that goes on in classrooms. We would encourage that this type of test item be included in international projects even though the negative factors of time and reliability are problems. When properly pre-tested for the purpose of establishing categories for coding, they provide information on both the content and processes of science which exceeds that available from multiple-choice formats alone.

As alternative evaluation methods are being developed for large international assessment projects in mathematics and science, we hope that this information will have direct relevance to classroom evaluation. Variation in evaluation methods which encourages diagnostic evaluation will be an important tool for the classroom mathematics and science teacher.

References:

Angell, C og Lie, S., 1990, **Fysikkeksamen og eksamensfysikk.** nr 4 i skriftserien fra SLS, Universitetet i Oslo.

Assessment of performance Unit (APU), 1981 - 1989, Several Reports, Department of Education and Science, Her Majesty's Stationery Office

Bloom, 1956, **Taxonomy of Educational Objectives.** Handbook I: Cognitive Domain. New York, David McKay.

Brekke, Kjærnsli, Lie, Gisselberg, Wester-Wedman, Prien, Weng, 1992, **The TIMSS Pre-Pilot Test; Experience Critical Comments and Recommendations from Norway, Sweden, Denmark**, TIMSS Report

Comber, L.C. og Keeves, J.P., 1973, **Science Education in nineteen countries.** New Yourk: John Wiley

Elevtenkande och Kurskrav i Naturvitenskaplig undervisning, EKNA, 1979-1989, Several Reports from EKNA-project

Harsfjord, V. og Dalin, P., 1988, **Læreren og naturfagundervisningen**, Report nr. 2 from The Norwegian SISS-project: The second International Science Study,Universitetsforlaget, Oslo

Johnson, S. 1987, *Assessment in Science and Technology*, **Studies in Science Education**, 14 (1987), 83-108

Kempa, R., 1986, **Assessment in Science**, Cambridge University Press, ISBN 0521278635

Rosier and Keeves, 1992, **The IEA Study of Science I: Science Education and Curricula in Twenty-Three Countries**, IEA volum 8, Pergamon Press

SAPA, 1967, **Science - A Process Approach**, Washington, The American Association of Science

Science 5/13, 1972, **With Objectives in Mind**, London, McDonald Educational

SCIS, 1974, **Teachers' Handbook**, Berkely, Lawrence Hall of Science

Tamir, P., 1987, *Science Practiacal Process Skills of Ninth grade Students in Israel*, Adey, Bliss, Head, Shayer, ed., 1987, Falmer Press

Tamir, P. 1990, *Justifying the selection of answers in multiple choice items*, **International Journal of Science Education**, vol 12, no 5

The Third International Mathematics and Science Study, 1991, **Project Overview**, ICC 200